DOCUMENT RESUME

ED 330 715 TM 016 297

AUTHOR Perlman, Carole L.
TITLE Developing a Comprehensive Assessment Program.
PUB DATE Apr 91
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Classroom Techniques; Educational Assessment; *Educational Change; Elementary Secondary Education; Inservice Teacher Education; *Program Development; *School Districts; *Standardized Tests; Teacher Education; Testing Problems; *Testing Programs; Test Use
IDENTIFIERS Chicago Public Schools IL

ABSTRACT

The use of standardized tests within the Chicago (Illinois) Public School System and elsewhere is discussed in the context of school reform efforts. Standardized tests have been criticized as: being too general to assess all valued outcomes; being too general to fit with local curricula; and directing instruction only to test objectives. Recent assessment designs have: relied more heavily on written, as opposed to multiple-choice, formats; increased the variety of performance tasks covered; and incorporated student portfolios. A study by R. J. Stiggins (1988) of school assessment indicated that: (1) teachers spend as much as 20 to 30% of their professional time directly involved in assessment-related activities; (2) teacher training in assessment does not aid teachers' day-to-day classroom activities; and (3) inadequate classroom assessment mat adversely affect learners. Based on this discussion, it is suggested that school districts continue implementing standardized testing programs, take steps to maximize the validity of assessments, use valid systemwide performance assessment to the greatest extent possible, promote high-quality classroom assessment, and use assessment results intelligently and appropriately. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED330715

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

CAROLE L. PERLMAN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Developing a Comprehensive Assessment Program

BEST COPY AVAILABLE

Carole L. Perlman
Chicago Public Schools

Paper presented at the annual meeting
of the American Educational Research Association
April 7, 1991
Chicago, Illinois

TM016297

"...educational reform in school districts organized under this article shall be implemented in such manner that...by the conclusion of the 1993-1994 school year, at least 50% of all students--regardless of race, ethnicity, gender or income status--in each attendance center within the district score at or above the national norm on a standardized test...."

Illinois Public Act 85-1418

"Something has to be done about testing. We need better data sources than tests. I don't believe any test scores I see anymore--even SAT, because we teach toward them so much. Testing is only a good measure if you don't pay any attention to it."

--Freda Holley (1988)

When the Chicago school reform law was written, the Illinois legislature obviously viewed standardized test scores as the ultimate indicator of whether individual schools and the reform process as a whole are successful. Indeed, the law specified that nationally normed standardized tests would be the *only* criterion measure of academic achievement. Politicians, the business community and other members of the public see accountability as the key to reform. Test scores are considered to be the definitive measure of the success of the reform.

People say "We've got to get those test scores up!" much more often than they say "I hope that the students learn to read better." That language is revealing; their stated goal is higher test scores rather than higher achievement. But if the scores go up, can we in fact conclude that reform has been a success?

Standardized tests, especially high-stakes tests, have been the object of much criticism during recent years. Much of this criticism centers on the content and format of the tests and on a closely related issue, the effect of high stakes standardized tests on curriculum and instruction. There is also concern that results can be misleading and used inappropriately. Since those criticisms have been discussed in detail elsewhere (e.g., Cannell, 1988; Shepard, 1989; Meisels, 1989; Haney & Madaus, 1989; Shepard, 1991), I will give only a brief summary.

The content of standardized tests has been criticized for being too general, for not covering all valued outcomes and for not fitting precisely with local curriculum, and for being inappropriate for some groups of children. The objectives measured on standardized tests come from textbooks and local and state curriculum guides, with the intention that the tests be general enough to be used in many different places. They were never designed to measure every nuance of local curriculum, but they do measure, in a very broad sense, how well a given school is performing in third-grade math or fifth-grade reading, compared with others in the norming sample. If the tests very closely matched local curriculum, they would have to be very much longer and would become less useful normative measures. Many critics find fault with the multiple choice format, since it requires that the student merely identify a correct answer, rather than produce one. It is rarely the case that in "real life" we are provided a menu of possible answers and asked to select the best one.

Also many questions simply don't have a single right answer. There is also a popular assumption that standardized tests do not measure higher order thinking skills. But Worthen and Spandel (1991) write "The notion that multiple choice tests can tap only recall is a myth. In fact, the best multiple choice items can--and do--measure students' ability to analyze, synthesize information, make comparisons, draw inferences...." (p. 67).

There is also concern that instruction can become limited to those objectives measured on the tests using formats similar to standardized test items. As Lorrie Shepard (1991) points out, this practice can result in test scores going up without a corresponding real gain in achievements. Many are troubled that instructional time is diverted from subject area instruction to test-taking skills. These practices can also compromise the validity of the test by making it questionable how well students' performance could generalize to other types of items in the domain of interest. Albert Shanker (1988) summarized the problem succinctly:

> Since the reputation of a school, its principal, its teachers, and the school board and superintendent depends largely on standardized test scores, schools are devoting less time to reading real books, writing essays, and discussing current events and more and more time teaching kids strategies for filling in blanks and choosing the answers to multiple choice questions. This destroys much of the value of these tests, which only tell you something if they are an independent measure of what the student knows.

Other problems relate to the ways that standardized test scores are reported and used. Aging norms and easily misinterpreted grade-equivalent scales have

been widely criticized. All too often a single piece of fallible test data is used with arbitrarily set cut-off points to make important decisions that the tests were never designed to make.

Given all of the above, standardized test results aren't going to tell us everything we need to know as educators or as citizens. We undoubtedly need to be looking at other kinds of indicators.

In recent years there has been increasing sentiment in favor of redesigning assessments to more closely approximate the actual criterion behaviors of interest. The most common example of performance assessment is a writing test that involves an actual writing sample rather than multiple choice writing mechanics items, but many other types of performance assessment are possible. For example, Connecticut's state assessment has used a variety of performance tasks in which:

- business education students perform tasks comparable to those performed by entry level secretarial workers

- students design science experiments and use laboratory equipment

- foreign language students write letters and hold conversations with trained assessment staff.

Ideally, performance assessments should closely resemble good instruction that emphasizes higher order thinking skills.

Portfolios of student work can also be a valuable assessment tool. Besides producing a rich database for parents, school staff and students (especially if the portfolios contain work done over a period of several years), they can result in students developing a sense of responsibility for their work and an ability to critically evaluate their work.

Ultimately, if reform is going to work, there must be reform in the classroom, as well as in school governance. For that reason, we need to take a careful look at the way assessment is carried out in the classroom. Richard Stiggins and his colleagues have been studying this subject for a number of years. They have discovered the following (Stiggins, 1988):

- Teachers may spend as much as 20-30% of their professional time directly involved in assessment-related activities, the results of which are instrumental in determining pacing and methods of instruction, whether instructional objectives have been met and students' grades.

- Even if teachers have received preservice training in assessment (and many have not), that training was not perceived as very helpful, since

that training was not perceived as being helpful in day-to-day classroom activities. That's because measurement courses have traditionally focused on standardized tests and statistical concepts. "Daily assignments, performance assessments, tests that accompany textbooks, and oral questions,--all key strategies of classroom assessment--are virtually ignored" (p. 365).

- Although teachers would welcome relevant and useful training or assistance, inservice training in classroom assessment is rare, as is the availability of direct technical assistance in dealing with day-to-day issues of assessment in the classroom.

- "Every day teachers need to assess thinking skills in the classroom, but they have not been trained or prepared to do so. Teacher-developed paper-and-pencil quizzes provided by textbook publishers are currently dominated by questions that ask students to recall facts and information. Although instructional objectives and even instructional activities may seek to develop thinking skills, classroom assessments often fail to match these aspirations." (p. 365).

- Inadequate classroom assessment can have a variety of serious negative effects on learners.

It is evident that classroom assessment is a crucial, but too often ignored component of a systemwide assessment program. We can no longer afford to neglect it. It is also clear that an emphasis on "getting those Iowa scores up" to the exclusion of all else will have devastating costs and will undermine the effectiveness of school reform. There must be instead a comprehensive assessment program, in which the parts can function in a complementary way, exploiting the unique advantages of each. The following strategies can lead us to that type of program.

1. Continue the standardized testing program.

Given the earlier discussion of standardized testing, one might wonder why we shouldn't get rid of it altogether. However, standardized tests can give a good general picture of how our students perform compared to others elsewhere. Also, the public, particularly the business community, is accustomed to getting standardized test scores and the sudden absence of them would be construed as an evasion. Even if performance assessments were implemented, there would still be considerable pressure to have an objective test as a back-up to the subjective measures. Compared with performance assessments, standardized tests have broad content coverage and they are relatively quick and inexpensive to administer. Finally, administration of standardized tests is required for pupils in ESEA Chapter 1 and some other state- and federally-funded special programs.

For all of their faults--and admittedly there are many--we can't realistically ascribe to standardized tests the blame for everything that's wrong with our schools. If the tests disappeared tomorrow, there would still be unimaginative curricula and poor teaching that emphasized drill, rote memory, and ditto sheets. All of those problems existed before the rise of standardized testing.

2. *Take steps to maximize the validity of assessments.*

A number of procedures have been recommended by the NCME Task Force on Enhancing the Credibility of School Testing Programs. Briefly, these involve using the newest norms available, using new tests every year, maintaining stringent security procedures, and limiting the amount and nature of test preparation.

3. *Use valid systemwide performance assessment to the extent possible.*

Our system currently conducts performance assessment in listening, speaking and writing and we are in the process of developing performance assessments in science. Performance assessment can provide invaluable information, but a number of issues must be kept in mind:

- Performance tests are expensive and time-consuming to develop, administer and score. To maximize validity these tests should be scored (and sometimes administered as well) by someone other than the student's teacher. The costs can be staggering in a system as large as ours: $2-$3 per student for essay scoring (or about $200,000-300,000 for our state-required assessment at grades 3, 6, 8 and 11) and $100 per classroom science assessment kit plus 1-2 tester-days (at about $60 per day) per classroom-- and there are about 1,000 classrooms per grade level. In contrast, standardized test materials and scoring for all subject areas cost less than $2 per student. It is true that sampling would cut costs, but that's not always feasible; sometimes it's necessary to assess every student. In a time of budgetary constraints, it will be difficult, if not impossible, to obtain sufficient funding to run a quality performance assessment program.

- Test reliability is a serious problem and outside of writing, scoring procedures for performance tests at the elementary and secondary levels are not well-developed.

- Content coverage on performance assessments in many subject areas tends to be very narrow.

- Although advocates of performance assessment argue that teaching to the test will have only salutary results, that is certainly not the case. In fact, teaching to a performance assessment can have even worse effects on instruction than teaching to a standardized test if the same performance items are used over and over (a strong temptation, since the items are quite expensive to develop and materials can be costly). The New York science assessment has used the same items over a period of years and it is common knowledge among fourth grade teachers that the assessment will require students to build an electrical circuit and classify beans. Anecdotal evidence suggests that preparation for that test can involve instruction that is even more narrow than the kind that teaching to a standardized test might engender. Also, if a child is given specific instruction building an electrical circuit and testing for conductivity without understanding how or why the process works, the assessment no longer measures higher order thinking, but rote memory instead.

Performance assessment clearly has value, particularly in those subject areas that do not readily lend themselves to objective tests, but it needs to be subjected to critical scrutiny. For example, currently used science performance assessments appear to be more a device for curricular and instructional improvement than a measurement tool. While the goal of improving instruction is admirable, it is necessary to distinguish between

curricular reform and measurement. We cannot say that performance assessment is the solution to all measurement problems because it is intuitively appealing. After all, teaching and measuring isolated subskills once seemed intuitively appealing, too.

4. *Promote high-quality classroom assessment.*

A key element of Chicago's school reform is the development of curricula tailored to each school's needs. Since no systemwide assessment could ever measure everything that local schools choose to teach, school staff will need to develop assessments to measure attainment of their unique local objectives. Because preservice measurement training for teachers has been inadequate, inservice training needs to be provided so teachers can learn how to design quality assessments, how to assess higher-order thinking skills, how to design and use portfolios and other performance measures, and how to critically evaluate text-embedded tests. Because we feel that improving classroom assessment can have a profound impact on improving instruction, we have begun to offer teacher workshops patterned after those conducted by Stiggins (1990). The workshops have been very well-received.

5. *Use assessment results intelligently and appropriately.*

Unfortunately decision makers all too rarely recognize that test scores are not error-free and that tests must be validated for each purpose for which they are used. A stunning example of this is our former systemwide promotion policy, adopted in 1985 over the protests of the Department of Research and Evaluation. If narrowly interpreted, the policy required that any student scoring below the fifth stanine on the Iowa could be retained unless he or she had been retained before. The policy was scrapped during the 1989-90 school year, but by then considerable damage had already been done. Our dropout rate will reflect the effects of that policy for many years to come.

There are two lessons to be learned from this. The first is that there needs to be greater understanding on the part of school staff and the general public of the uses and limitations of tests results. The second lesson to be learned is that multiple measures must be used whenever possible, both in making decisions about children and about schools. In our case those might include Iowa scores, state assessment results, results of performance assessments, portfolios, grades, and teachers' ratings. In evaluating schools, it is necessary to look beyond test scores to the quality and types of educational experiences provided to the students: How many and what kinds of books do the children read? Do they often write compositions? Does "hands-on" math and science instruction take place? Is there instruction in the arts?

If the Iowa is no longer the sole measure of success, teaching it to the exclusion of all else will become an increasingly less attractive and less feasible option.

References

Cannell, J. J. (1988). Nationally normed elementary achievement testing in America's public schools: How all 50 states are testing above the national average. Educational Measurement, 7(2), 5-9.

Haney, W. & Madaus, G. (1989). Searching for alternatives to standardized tests: Whys, whats, and whithers. Phi Delta Kappan, 70(9), 683-687.

Holley, F. (1988, Winter). Interview with Freda Holley. Pre-Post Press, p. 7.

Meisels, S. J. (1989). High stakes testing. Educational Leadership, 46(7), 16-22.

Shanker, A. (1988, April 24). Exams fail the test. New York Times, E, p. 7.

Shepard, L. A. (1989). Why we need better assessments. Educational Leadership, 46(7), 4-9.

Shepard, L. A. (1991). Interview on assessment issues with Lorrie Shepard. Educational Researcher, 20(2), 1991.

Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest instructional priority. Phi Delta Kappan, 69(5).

Stiggins, R. J. (1990) Classroom Assessment Training Program. Portland, Ore.: Northwest Regional Educational Laboratory.

Worthen, B. R. & Spandel, V. (1991). Putting the standardized test debate in perspective. Educational Leadership, 48(5), 65-69.